Model Identification, Parameter Estimation, and Dynamic Flux Analysis of *E. coli* Central Metabolism

S. Čerić and Ž. Kurtanjek*
University of Zagreb, Faculty of Food Technology and Biotechnology
Pierottijeva 6, 10 000 Zagreb, Croatia

Original scientific paper
Received: March 20, 2006
Accepted: May 5, 2006

Dedicated to Prof. Dr. Đurđa Vasić-Rački on occasion of her 60th birthday

In this work are applied three global optimisation algorithms for adaptation of the mathematical model of the central metabolism of *Escherichia coli* to data obtained in the experiment with response to glucose impulse. Applied is the adaptive simplex method by Nelder-Mead, evolutionary algorithms of differential evolution, and simulated annealing. The original model has been modified by the following steps: closure of Entner-Doudoroff pathway with pyruvate balance, introduction of phosphoenolpyruavate carboxylase and carboxykinase reactions in the balance of phosphoenolypyravate, account for loss of pyruvate in biomass synthesis, change in kinetic rate expressions for several enzymes, and partial re-estimation of the kinetic parameters by the global optimisation algorithms. The modified model correctly predicts observed oscillatory response to glucose impulse in concentrations of pyruvate and D-ribose-5-phosphate. To discern metabolic control, evaluated are dynamic intracellular fluxes by the model simulation around the following network branching metabolites: α-D-glucose-6-phosphate, 6-phospho-D-gluconate, glyceraldehydes-3-phosphate, and pyruvate. The simulation of the fluxes around phosphoenolypyruvate show that phosphoenolpyruavate carboxylase and carboxykinase (PEPCK) activity and phosphotransferase system (PTS) are closely dynamically tied, indicating that glycolysis and TCA metabolisms can not be separated under the given transient conditions. Overall model adequacy is evaluated by standard deviations of the model predictions and experimental data for each metabolite.

Key words:
Escherichia coli, central metabolism, glucose impulse, dynamic metabolic flux analysis, global optimisation

Introduction

From industrial biotechnology view point, metabolic engineering is the improvement of cellular activities by manipulation of enzymatic transport and regulatory functions of the cell with the use of genetic engineering manipulations with purpose to improve the present and, more importantly, to develop new biotechnological processes. Availability of experimental techniques to manipulate genome greatly exceeds our present information and understanding of complexity of cell regulation. In order to have a rational approach in application of possible genetic manipulations, a systematic study of the system by means of metabolic engineering rules is needed.[1–2] Development of mathematical models of cellular metabolism is needed to guide planning of experimental genetic engineering manipulations. Due to availability of genomic data of many industrially important microorganisms, metabolic mathematical models are based on complete information on details of reaction networks. The most immediate application of genomic data in mathematical modelling are the structural metabolic models expressed in a form of complete stoichiometric matrices including all possible reactions coded in genome. The stoichiometric matrix is used as a set of mass balance constraints which need to be complemented with experimental data of extracellular rates, and/or thermodynamic feasibility of reactions, with possible inclusion of cybernetic principles for purposeful cell adaptation. From industrial stand point, cell adaptation is strongly dependent on numerous extracellular (reactor) conditions, such as composition of synthetic or natural nutrients, temperature, pH, hydrodynamics, viscous shear stress, and others. In order to develop metabolic models under realistic unsteady state bioreactor conditions, experimental data on intracellular activity are needed under variate working conditions.

Important attempts to cope with the metabolism regulation are models that include, at kinetic level, the main fluxes of the central metabolism (phosphotransferase system, glycolosis, penthose-phosphate, precursors).[3–7] These works provide es-

*Corresponding author: E-mail: zkurt@pbf.hr

sentially new experimental information on response of intracellular concentrations of *Escherichia coli* under transient conditions imposed by Dirac or Heaviside glucose impulse in a laboratory bioreactor. The data are obtained with very sophisticated and fully automated sampling and analytical system (based on LC-MS/MS). Intracellular composition is sampled at the frequency of 4 Hz (i.e. in intervals of 0.25 s). Due to a short time window of observations (about 20 s) gene regulatory effects do not have an impact on the observations, and only complex feedforward and/or feedback interaction of metabolites and coenzymes with enzymes are observed. Samples are immediately frozen and deactivated, and due to high sampling rate, data are believed to capture essential dynamic effects of enzyme kinetics and regulation. The data are average values of cells from samples drawn from bioreactor at intensive mixing rate to avoid spatial distribution of cells with different metabolic states.

This work is aiming to use the experimental data of *D. Degenring*,[6] with glucose impulse fed to *Escherichia coli* K12, to test global optimisation techniques for estimating the kinetic parameters, and to improve an extended kinetic model for the *E. coli* metabolism. Especially, predictions of the experimentally observed complex oscillatory behaviour in pyruvate and D-ribose-5-phosphate (which the original model does not predict) will provide a "strong" test for model validation. Based on the improved models a dynamic metabolic flux analysis around the key branching metabolites will be determined.

Modelling

The original model, adopted from literature and reduced by using the parameter sensitivity and the principal component analysis, is a stiff system of 10 highly nonlinear ordinary differential equations including 24 expressions for the kinetic rates.[6,7] The model equations can be written in the following matriceal form:

$$\alpha = \begin{pmatrix} +1 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 \\ -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & +1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & +1 & +1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 \\ 0 & +2 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & +2 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -6 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{r} = \begin{pmatrix} v_{\mathrm{PTS}} \\ v_{\mathrm{PGI}} \\ v_{\mathrm{PFK}} \\ v_{\mathrm{FBPuse}} \\ v_{\mathrm{ALDO}} \\ v_{\mathrm{TIM}} \\ v_{\mathrm{GAPPEP}} \\ v_{\mathrm{PK}} \\ v_{\mathrm{PDH}} \\ v_{\mathrm{G6PDH}} \\ v_{\mathrm{G3PDH}} \\ v_{\mathrm{PGDH}} \\ v_{\mathrm{TKATA}} \\ v_{\mathrm{PGHL}} \\ v_{\mathrm{GLP}} \\ v_{\mathrm{MUR}} \\ v_{\mathrm{FBPBM}} \\ v_{\mathrm{GLY}} \\ v_{\mathrm{PEPBM}} \\ v_{\mathrm{SER}} \\ v_{\mathrm{RPPK}} \\ v_{\mathrm{DAHP}} \\ v_{\mathrm{PEPCK}} \\ v_{\mathrm{PYRBM}} \end{pmatrix} \tag{1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{c} = \alpha^T \cdot \boldsymbol{r}(\boldsymbol{c},\boldsymbol{c}_u,\beta) = \begin{pmatrix} c_{\mathrm{G6P}} \\ c_{\mathrm{F6P}} \\ c_{\mathrm{FBP}} \\ c_{\mathrm{GAP}} \\ c_{\mathrm{DHAP}} \\ c_{\mathrm{G3P}} \\ c_{\mathrm{PEP}} \\ c_{\mathrm{PYR}} \\ c_{\mathrm{6PG}} \\ c_{\mathrm{C5P}} \end{pmatrix} \tag{2}$$

Here the kinetic functions of the original model have been modified by the following functions:[8]

In Fig. 1. the reaction network is graphically presented, by including the model reactions, individual metabolites, and pools of metabolites introduced after the model reduction. Balances for the cofactors (ATP, ADP, NADH, NAD) are not accounted for explicitly in the model, being included into the reaction kinetic as experimentally determined time varying interpolation functions.[6,7]

Due to the complex interconnectivity between the flux balances, and the wide range of time scales in the enzyme kinetics, integration of the model equations is subjected to possible numerical errors.[8,9] In the present study, a very accurate and reliable self adaptive integration procedure NDSolve, provided by Wolfram Reasearch *Mathematica,*[10] has been applied.

for phosphotransferase system

$$v_{\mathrm{PTS}} = \frac{V_{\mathrm{fPTS}} \cdot c_{\mathrm{GLC}} \cdot c_{\mathrm{PEP}}}{(K_{\mathrm{mGLCpts}} + c_{\mathrm{GLC}}) \cdot (K_{\mathrm{mPEPpts}} + c_{\mathrm{PEP}})} \cdot \frac{1}{(K_{\mathrm{iG6Ppts}} + c_{\mathrm{G6P}})^{\mathrm{npts}}} \tag{3}$$

glucose-6-phospho isomerase

$$v_{\mathrm{PGI}} = \frac{V_{\mathrm{fPGI}} \cdot \frac{c_{\mathrm{G6P}}}{K_{\mathrm{mG6Ppgi}}} - V_{\mathrm{bpgi}} \cdot \frac{c_{\mathrm{F6P}}}{K_{\mathrm{mF6Ppgi}}}}{1 + \left(\frac{c_{\mathrm{F6P}}}{K_{\mathrm{iF6Ppgi}}}\right) + \left(\frac{c_{\mathrm{G6P}}}{K_{\mathrm{mG6Ppgi}}}\right) + \left(\frac{c_{\mathrm{F6P}}}{K_{\mathrm{mF6Ppgi}}}\right) \cdot \left(1 + \frac{c_{\mathrm{F6P}}}{K_{\mathrm{iF6Ppgi}}}\right)} \tag{4}$$

aldolase

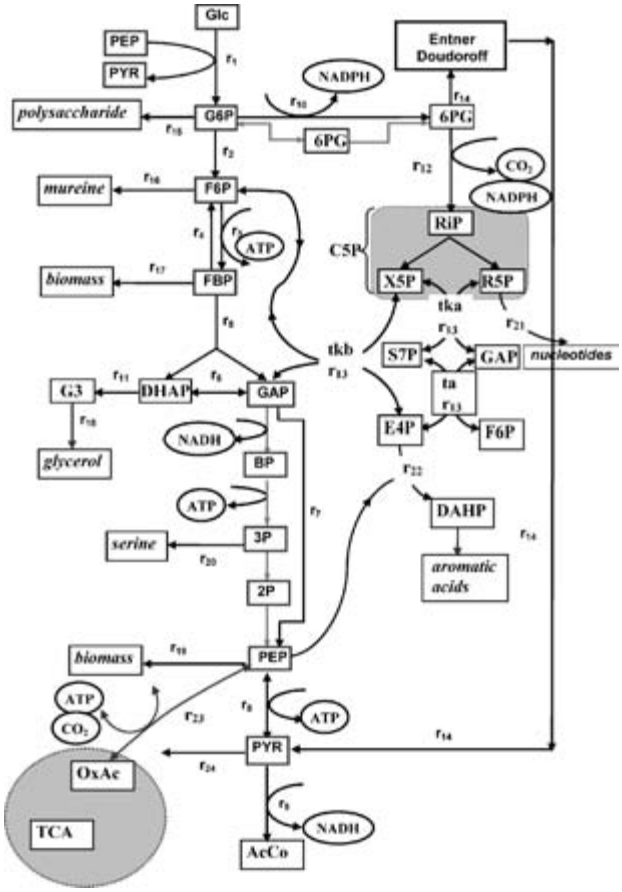
$$v_{\mathrm{ALDO}} = \frac{\left(V_{\mathrm{max\,aldo}} \cdot \left(c_{\mathrm{FBP}} - \frac{c_{\mathrm{G3P}} \cdot c_{\mathrm{DHAP}}}{K_{\mathrm{eqaldo}}}\right) \cdot \left(1 + \frac{c_{\mathrm{F6P}}}{K_{\mathrm{aF6Paldo}}}\right)\right)}{\left(\begin{array}{l} K_{\mathrm{mFBPaldo}} + c_{\mathrm{FBP}} \cdot \left(1 + \frac{c_{\mathrm{G3P}}}{K_{\mathrm{iG3Paldo}}}\right) + \frac{V_{\mathrm{max\,aldo}} \cdot \left(1 + \frac{c_{\mathrm{F6P}}}{K_{\mathrm{aF6Paldo}}}\right)^n}{V_{\mathrm{max\,aldo}} \cdot K_{\mathrm{eqaldo}}} \cdot \\ \cdot (K_{\mathrm{mDHAPaldo}} \cdot c_{\mathrm{G3P}} + K_{\mathrm{mDHAPaldo}} \cdot c_{\mathrm{DHAP}} + c_{\mathrm{G3}} \cdot c_{\mathrm{DHAP}} \end{array}\right)} \tag{5}$$

The new adopted kinetic expressions are selected from the literature,[6] by using numerous model simulations with the aim to achieve a better fit with the experimental data.[8] Additional rates are:

$$v_{\mathrm{PEPCK}} = \frac{V_{\mathrm{fpepck}} \cdot \frac{c_{\mathrm{PEP}}}{K_{\mathrm{mPEPpepck}}} - V_{\mathrm{bpepck}} \cdot \frac{c_{\mathrm{OAA}}}{K_{\mathrm{mOAApepck}}}}{1 + \left(\frac{c_{\mathrm{PEP}}}{K_{\mathrm{mPEPpepck}}}\right) + \left(\frac{c_{\mathrm{OAA}}}{K_{\mathrm{mOAApepck}}}\right)} \tag{6}$$

and the loss of pyruvate during the biomass synthesis:

$$v_{\mathrm{PYRBM}} = \frac{V_{\mathrm{fPyrBM}} \cdot c_{\mathrm{PYR}}^{\mathrm{nPyrBM}}}{K_{\mathrm{mPyrBM}}^{\mathrm{nPyrBM}} + c_{\mathrm{PYR}}^{\mathrm{nPyrBM}}} \tag{7}$$



F i g . 1 – *Reaction network of the glycolysis, penthose phosphate and Entner-Dourdoff pathways of Eschericchia coli K 12. The reactions labels are:* $r_1 = v_{PTS}$, $r_2 = v_{PGI}$, $r_3 = v_{PFK}$, $r_4 = v_{FBPase}$, $r_5 = v_{ALDO}$, $r_6 = v_{TIM}$, $r_7 = v_{GAPPEP}$, $r_8 = v_{PK}$, $r_9 = v_{PDH}$, $r_{10} = v_{G6PDH}$, $r_{11} = v_{G3PDH}$, $r_{12} = v_{PGDH}$, $r_{13} = v_{TKATA}$, $r_{14} = v_{PGHL}$, $r_{15} = v_{GLP}$, $r_{16} = v_{MUR}$, $r_{17} = v_{FBPBM}$, $r_{18} = v_{GLY}$, $r_{19} = v_{PEPBM}$, $r_{20} = v_{SER}$, $r_{21} = v_{RPPK}$, $r_{22} = v_{DAHPS}$, $r_{23} = v_{PEPCK}$, $r_{24} = v_{PYRBM}$.

Model parameter estimation

The model properties essentially depend on the numerous kinetic parameters, which determine the reaction rates, saturation constants, and strong activation and inhibition effects of numerous cofactors and metabolites distributed throughout the network. Thus, even if difficult, the parameter estimation is an essential numerical step in analysis. This step can be performed by minimizing the unweighted sum of squares of residuals, i.e. the errors between the model predictions of the metabolite concentrations and measured data, given by:

$$S^2(\beta) = (\boldsymbol{c} - \boldsymbol{c}_{\mathrm{m}}) \cdot (\boldsymbol{c} - \boldsymbol{c}_{\mathrm{m}})^T \qquad (8)$$

In this work, three potentially effective algorithms for global optimisation have been comparatively tested in order to search for the kinetic parameter of the *E. coli* central metabolism model.

Nelder-Mead optimisation

The first applied search algorithm is the simple, but the very "code cost effective" method of Nelder-Mead (NM)[11]. This method can be described as a "down hill simplex descent". Although it is an "old" method, it has a "reputation" of being effective if a suitable initial parameter guess is provided. The search starts with an initial (usually randomly picked) simplex of $N+1$ vertices in the N dimensional space. The simplex edges are taken as unit vectors, and at the vertices objective function is evaluated, and values sorted in decreasing order. In the next step the vertex with the highest value is improved. Centroid of N vertices is calculated, excluding the highest (labelled with $i = 1$), by:

$$\boldsymbol{x}_{\mathrm{mean}} = \frac{1}{N} \cdot \sum_{i=2}^{N+1} \boldsymbol{x}_i \qquad (9)$$

From this point of high f-value, a new search direction is determined by reflection from the centroid

$$\boldsymbol{x}_1^{\mathrm{new}} = \boldsymbol{x}_{\mathrm{mean}} + (\boldsymbol{x}_{\mathrm{mean}} - \boldsymbol{x}_1) \qquad (10)$$

If $f(\boldsymbol{x}_1^{\mathrm{new}}) < f(\boldsymbol{x}_{N+1})$ then the new vertex is in a downhill direction. A better point is attempted by doubling the move. If after the reflection the new point is still the highest, a reflection and shrinking is attempted:

$$\boldsymbol{x}_1^{\mathrm{new}} = \boldsymbol{x}_{\mathrm{mean}} + \frac{1}{2} \cdot (\boldsymbol{x}_{\mathrm{mean}} - \boldsymbol{x}_1) \qquad (11)$$

If this does not improve, then just shrinking is attempted:

$$\boldsymbol{x}_1^{\mathrm{new}} = \boldsymbol{x}_{\mathrm{mean}} - \frac{1}{2} \cdot (\boldsymbol{x}_{\mathrm{mean}} - \boldsymbol{x}_1) \qquad (12)$$

Finally, if this also fails, then all the vertices are shrunk toward the "best" one

$$\boldsymbol{x}_1^{\mathrm{new}} = \boldsymbol{x}_i - \frac{1}{2} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_{\mathrm{N+11}}) \quad i = 1,2,\ldots,N \qquad (13)$$

Through successive iterations, a simplex tumbles downhill with a continuous change of scale and adaptation in size and location following the configuration of the response surface. However, for an efficient and successfull localization of the global minimum, a good starting simplex is required, together with restart of search by using different initial simplexes.

Simulated annealing

Simulated annealing (SA) is inspired from thermodynamic principles on energy distribution in

multicomponent systems.[9] The name of the algorithm derives from the analogy between simulation of the annealing process of solids. Starting from an initial state (initial value of the model parameters), the system is perturbed at random to a new state in the neighbourhood of the original one, for which a change of ΔE in the objective function is evaluated. In a minimization process if the change ΔE is negative then the new state (model parameters) are accepted. Following the statistical thermodynamics, a probability p to find a system in a state with energy E at temperature T is given by the Boltzman law:

$$p(E,T) \propto \mathrm{e}^{-E/kT} \tag{14}$$

where k is the Boltzman constant. The value of an objective function f (sum of squares of the model residuals) is associated with the energy E of a system. Energy of a system state is calculated starting from an initial set of randomly picked points and an initial system temperature. A new state, reachable from the current state, is randomly selected and its energy is evaluated. If this energy is lower, the new state is always adopted. But if it is higher, the new state is accepted based on a fluctuation probability of size $exp(-\Delta E/kT)$, which must be smaller than with an uniformly distributed random number from the interval [0,1]. Successively, "temperature" of the system is lowered and smaller energy fluctuations become more statistically significant. Theoretically, as $T \to 0$, the system energy approaches its global minimum. However, in practice, the search is stopped when a maximum number of iterations exceed a predefined limit. Restarting the procedure with a new set of initial points increases a chance to reach the global minimum.

Differential evolution

The differential evolution (DE) method is a variant of the genetic algorithm (GA). It starts with a population of n random vectors $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_n$ of real numbers of N-dimension, called "genes". The initial vector population should be chosen randomly and should cover the entire search space (range of values of the model parameters). Compared to Nelder and Mead method, the initial population for DE should provide a better chance for the method to converge to global optima. In every iteration, for each $\boldsymbol{x}_i$ are chosen random integers a_1, a_2, and a_3 by which is constructed a corresponding mate

$$\boldsymbol{x}_i = \boldsymbol{x}_{a1} + \gamma \cdot (\boldsymbol{x}_{a2} - \boldsymbol{x}_{a3}) \tag{15}$$

where γ is fixed and predetermined scaling factor. Then $\boldsymbol{x}_i$ is mated with $\boldsymbol{y}_i$ according to the given crossover probability. Gene exchange is performed by exchange of vector components. In addition, a point mutation randomly occurs at randomly selected component. The result of a mutation is a child vector $\boldsymbol{z}_i$, which competes in terms of by fitness function with its parent $\boldsymbol{x}_i$ for a place in the new search population.

The DE-GA method is very robust, but to increase the convergence toward the global minimum, iterations must be restarted from several initial populations.

Results and discussion

The kinetic model (2) estimation results are presented in Fig. 2–3. The objective function (8) minimization has been performed by using the all three mentioned optimization methods, but the results by DE are accepted as the most probable. Initial guess points are randomly selected by the algorithms provided by Wolfram Research "*Mathematica*"[10]. In Fig. 4 are depicted experimental values and the model predictions for concentrations of the metabolite. Results of the dynamic flux analysis around the branching points of the reaction network are graphically presented in Fig. 5a-e.

Parameter estimation

A typical example of the search trajectory during the objective function minimization by the three methods is given in Fig. 2. The "landscape" of the objective surface (in the parameter space) depends on the selected model, but also on the level of "noise" present in experimental data. Due to numerous parameters, it is expected that the surface will present many local minima which would make minimisation by a gradient based method ineffective, i.e. it would have to been restarted many times in order to reach an acceptable minima. In contrast, when the initial points are randomly selected, and since the random search procedures do not depend

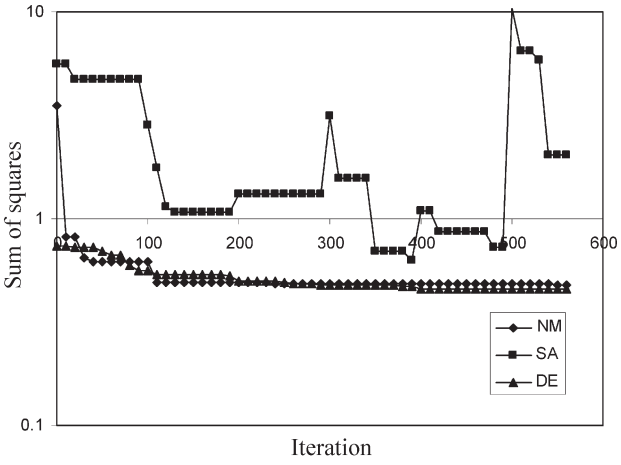


Fig. 2 – *Minimization of the sum of squares of residuals by Nelder-Mead (NM), differential evaluation (DE), and simulated annealing (SA) algorithms*

on the gradient, there is a better chance to reach the "global" minimum. The results, presented in Fig. 2 for NM and DE-GA, show that the two methods reach almost the same value of minima of the objective function, of approximately the same rate of convergence. The simulated annealing (SA) does not approach the minimum in a monotonic fashion, and consequently is less effective. Although NM and DE methods indicate almost the same minimization results, when they are compared in the parameter space some differences can be observed. In Fig. 3a–b are given the projections of the search movements on the plane of maximum velocity v_{fpTS} and saturation constant with respect to glucose K_{mGlcPTS} in the PTS model. The optimal parameter values are different, and parameter search trajectories are also different. The MN method initially presents a high convergence rate and large downhill steps, but its convergence becomes very slow near the minimum. In this study, the DE method presents better search properties since it

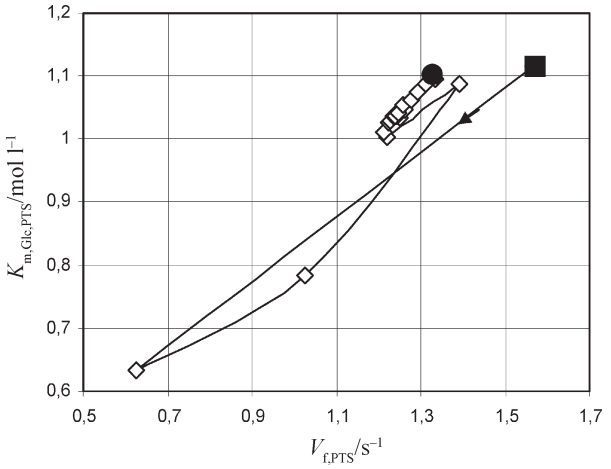


Fig. 3a – *Tracking of phosphotransferase parameters during minimization by Nelder-Mead algorithm. The initial point is denoted with* ■*, and the final point with* ●

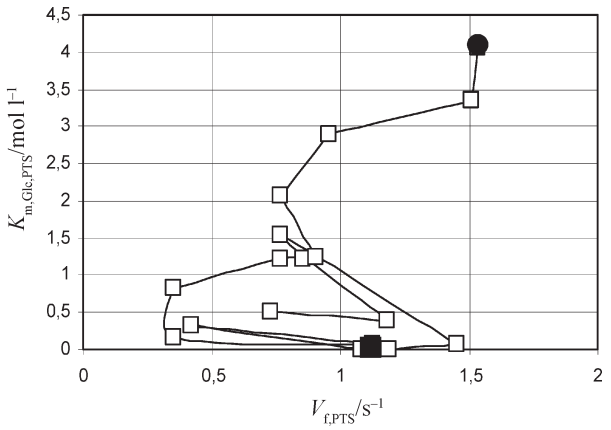


Fig. 3b – *Tracking of phosphotransferase parameters during minimization by the differential evolution algorithm. The initial point is denoted with* ■ *and the final point with* ●

"covers" a wider area in the parameter space, and its convergence rate does not slow down near the minimum. The fact that the parameters do not converge to the same values can be explained by the well known difficulties to distinguish among separate effects of the maximum rate and the saturation constant in the Michaelis-Menten kinetic rate expressions.[9] Statistically, for a given confidence level, the joint confidence region of these two parameters looks like an elongated ellipse. From practical reasons, the inability to do precisely with accuracy such high correlated parameters is not crucial since in this work the modelling objective is to obtain flux (adequate reaction rate) predictions which are less sensitive to parameter variations inside the joint confidence interval.

Metabolite responses

Quality of the model predictions vs. the experimental data, by using the estimated parameters from Table 1, can be observed in Fig. 4. The "steady state" data prior to the introduction of the impulse are not modelled due to inability to apply steady state flux analysis without reliable experimental data for glucose influx determined by its consumption within the reactor. In general, the model predictions of the metabolite concentrations under the transient conditions are in a good agreement with the experimental data, standard deviations between the model and experiments being given in Table 2. The average standard deviation is of 0.4 mmol L^{-1}, with a maximum value of 0.71 for DHAP and a minimum of 0.1 mmol L^{-1} for G6P. Most of the data points seem to be randomly distributed around

Table 1 – *Parameter estimates obtained from various starting vales.[8] The saturation constants are given in mmol L^{-1} and the reaction rates in mmol L^{-1} s^{-1}.*

v_{fpts}	1.336	$K_{iAcCoApdh}$	1.95689
K_{mGLCpts}	1.09438	*npdh*	1.60881
K_{mPEPpts}	0.006718	v_{fg6pdh}	2.095
K_{iG6Ppts}	0.43089	v_{fpgdh}	8.56
npts	3.7678	K_{eqpgdh}	10.5
v_{ftim}	4.636677	v_{ftkata}	0.01134
K_{mDHAPtim}	4.92799	v_{frppk}	0.0616
v_{btim}	3.69274	*nrppk1*	0.1477
K_{mGAPtim}	4.06253	$K_{mC5Prppk}$	2.346
v_{fgly}	1.32615	$K_{aADPrppk}$	0.14891
v_{fpdh}	6.3344	*nrppk2*	4.526
K_{mPYRpdh}	1.64019	v_{fdahps}	0.04
K_{mNADpdh}	0.104662	v_{bpepck}	0.7

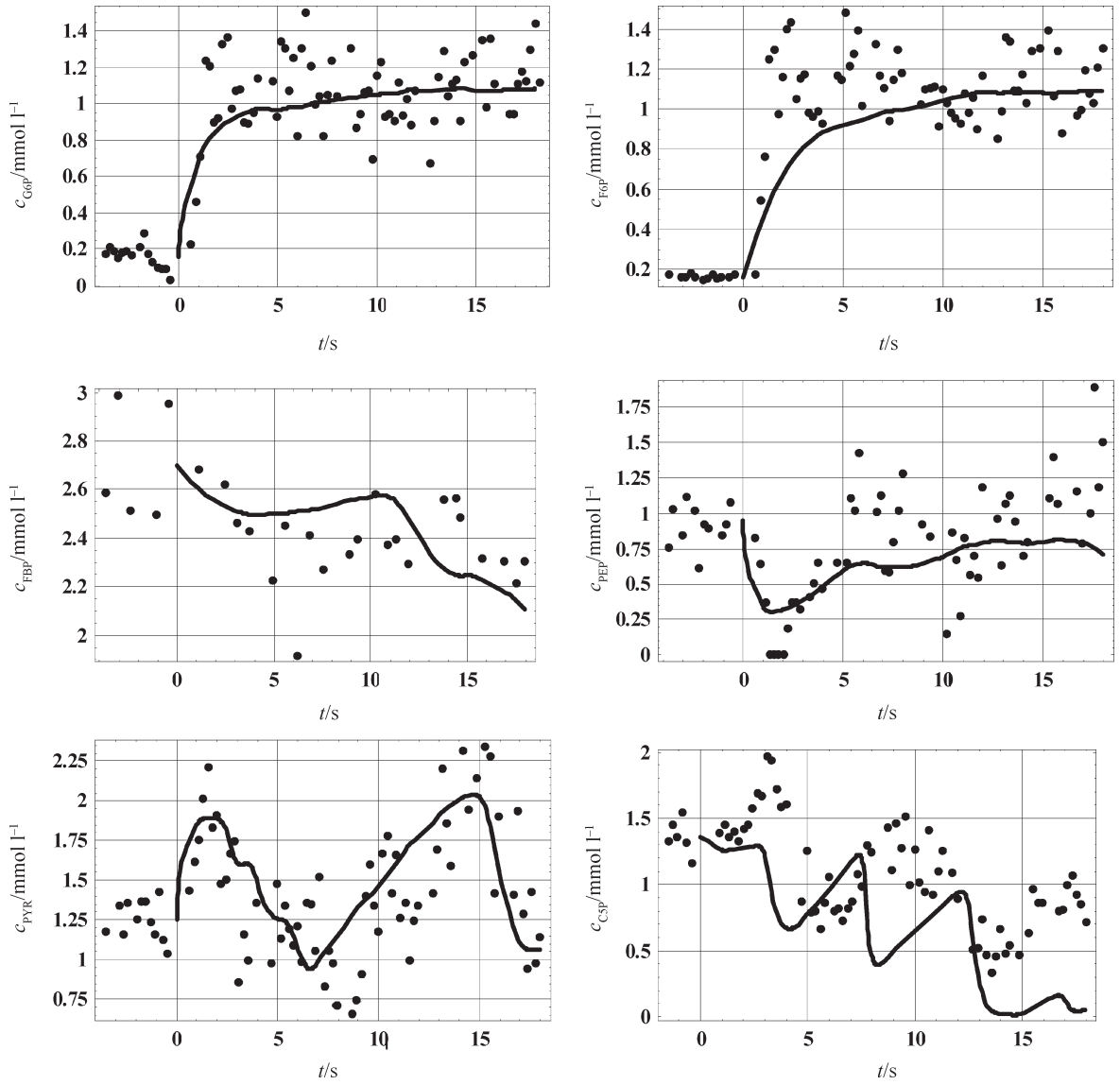


Fig. 4 – *Dynamic responses of intracellular metabolite concentration. Experimental data are shown as full circles ●, and the model predictions with continuous curves –.*

Table 2 – *Standard deviations between the measured intracellular concentrations and the model predictions under impulse induced transient conditions*

	σ / (mmol L^{-1})
G6P	0.11
F6P	0.21
FBP	0.19
GAP	0.60
DHAP	0.71
G3P	0.25
PEP	0.31
PYR	0.31
6PG	0.83
C5P	0.55

the model predictions (see G6P, F6P, PYR, PEP). For FBP, the model correctly predicts the decrease of the initial concentration, but outliers in experimental data seems to be present. In the case of C5P, it is to note a phase shifting between data and model, but the trend is correctly predicted. It is difficult to estimate the "true" model error due to the lack of an estimate for the measurement error under the working conditions. The analytical error under transient conditions should account for nonlinear interference of all metabolites and enzymes in the analysed sample with examined analyte. This error is assumed to be significantly higher compared to the analytical error in samples containing standard purified metabolites/compounds.

The main improvement in model predictions, comparatively to the previously published results of *Degenring* et. al.,[7] is better approximation of the

model response in the case of PEP, PYR and C5P. The main discrepancy between experimental data and the Degenring model predictions is present in the case of PEP concentration, that is a complete depletion of PEP, which is in opposition to the experimental results. Such a model prediction implies that PEP exercises dominant effect in PTS regulation, which is not supported by the experimental data. The wrong behaviour has been corrected by adding the PEPCK reactions for immediate replenishment of PEP by TCA reactions, and by introduction of the inhibitory effect of G6P on PTS.

This model correction is particularly important for predictions of PYR, due to its dominant position in the main flux distribution throughout the reaction network. In the work of *Degenring* et. al.,[7] their model predicts the first inflow followed by efflux of the pool, but fails to predict the secondary wave. In contrast, the modified model predicts oscillatory response in PYR pool, and also corrects the prediction of the oscillatory response of the C5 pool.

Dynamic flux analysis

The dynamic fluxes over the pathway are calculated around the following branching nodes (metabolites): G6P, 6PG, GAP, PYR, and PEP (see Fig. 5.a-e). Analysis of the flux dynamics can reveal insights into regulation kinetics. In Fig. 5a are shown the positive inlet flux r_{PTS} and the negative efflux of r_{PGI}, $r_{\mathrm{GLP,}}$ and r_{G6PDH}. The intensive influx of glucose is followed by an immediate response, leading to an increase of efflux by glucose-6-phosphate isomerase r_{PGI} to F6P pool. There are no immediate responses that can indicate an increase in the synthesis of polysaccharides and toward pentose phosphate pathway, these keeping at relatively low levels. However, the increase in the activity of isomerase can not compensate PTS and the unbalanced results suggested by the significant accumulation of G6P. This observation indicates a strong inhibitory effect of the accumulated G6P on PTS. An apparent steady state is established for about 10 s after the impulse perturbation, when the inlet PTS flux is balanced by the sum of the efflux flows.

In Fig. 5b. are displayed the fluxes around the 6PG pool. The overall dynamics is relatively slower when compared to the G6P pool. The inlet flux r_{G6PDH} slowly increases after the glucose impulse, but the efflux flux through Entner-Doudoroff r_{PGHL} and r_{PGDH} pentose phosphate pathways increase unproportionally, diminishing and disappearing in about 15 s, while the flow through PP pathway decreases and reaches a constant vale. As the net result of the differences, a constant decrease in 6PG pool concentration is observed, as confirmed by the experimental data.

In Fig. 5c the fluxes around GAP after the glucose impulse are presented. The inlet flux r_{ALDO} from FBP presents positive values, with an initial slow decrease, but then (after 10 s) presenting a significant increase. The flux r_{TIM} between DAHAP and GAP presents an alternating behaviour: firstly it flows toward DAHP, but after 10 s it reverses toward GAP. The efflux r_{GAPPEP} to PEP is apparently

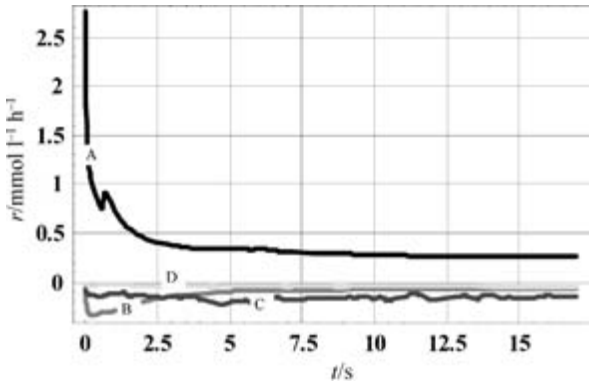


Fig. 5a – *Simulation of the fluxes around G6P after the glucose impulse. The inlet flux r_{PTS} (A-black) has positive values, and the efflux flow r_{PGI} (B-red), r_{GLP}(C-blue), r_{G6PDH} (D-green) have negative values.*

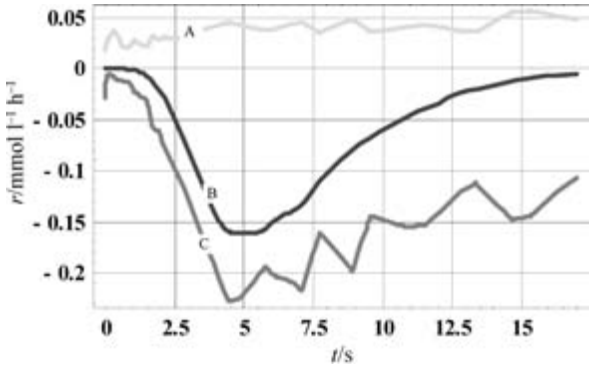


Fig. 5b – *Simulation of the fluxes around 6PG after the glucose impulse. The inlet flux r_{G6PDH} (A-green) has positive values, and the efflux flow r_{PGDH} (C-red), r_{PGHL} (B-blue) have negative values.*

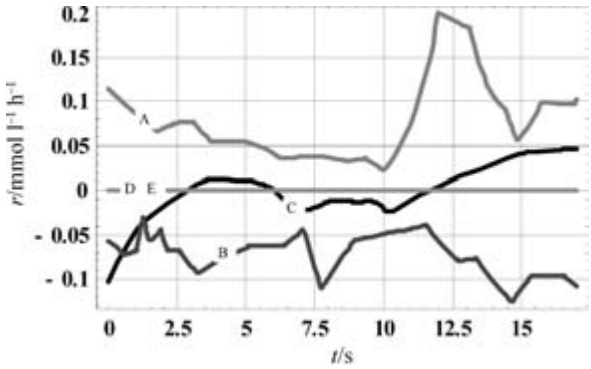
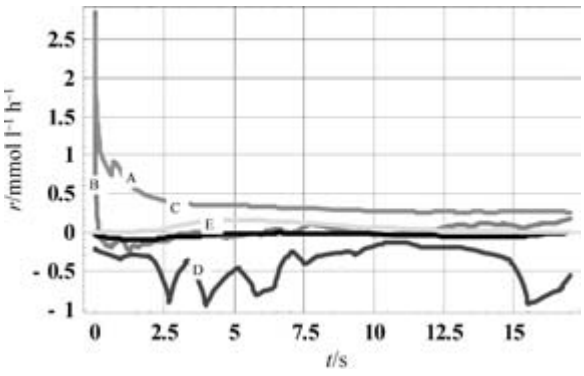


Fig. 5c – *Simulation of the fluxes around GAP after the glucose impulse. The inlet flux r_{ALDO} is denoted as (A red) has positive values, and the efflux flow r_{GAPPEP} (B blue), r_{TIM} (C black) have alternating values (directions), and the fluxes r_{SER} and r_{TKATA} (D and E in purple) have relatively negligible values.*

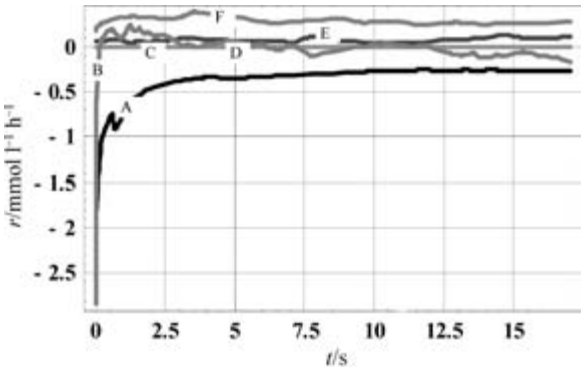
constant throughout the whole period, while the fluxes toward serine synthesis r_{SER} and backflow r_{TKATA} to F6P are relatively negligible.

Dynamic distribution of the fluxes around the key metabolite PYR is displayed in Fig. 5d. Initially, the high influx r_{PTS} to pyruvate by PTS is closely followed by the activation of the pyruvate kinase, r_{PK}. However, the flow from PEP to PYR quickly decreases to zero (during the first second), as PEP decreases due to a backward reverse of the PTS activity. The contribution to pyruvate via Entner-Doudoroff pathway is relatively negligible. Efflux r_{PDH} to AcCoA is irreversible but the activity of pyruvate dehidrogenase is time varying due to the cofactor NAD. The flux from pyruvate to biomass r_{PYRBM} is relatively negligible compared to the main fluxes.

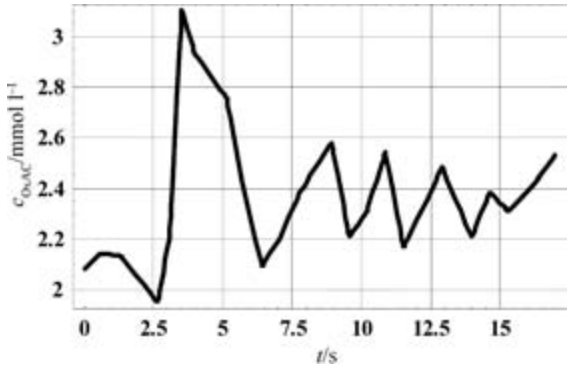
The balance of PEP is tied to the PTS mechanism and, through PEPCK reactions, to the TCA cycle, thus exemplifying during the short transient conditions the "collective" and concentrated type of regulation of the primary PTS mechanism. The simulations of the balance dynamics are presented in Fig. 5e, while the experimental data for oxaloacete in Fig. 6. Consumption of PEP is compensated by



F i g . 5 d – *Simulation of the fluxes around pyruvate after the glucose impulse. The inlet fluxes r_{PTS} (A purple) and r_{PGHL} (E-green) have positive values, and the efflux flow r_{PDH} (D–blue) and r_{PYRBM} (E-black) have negative values, while r_{PK} (B-red) has alternating values (directions).*



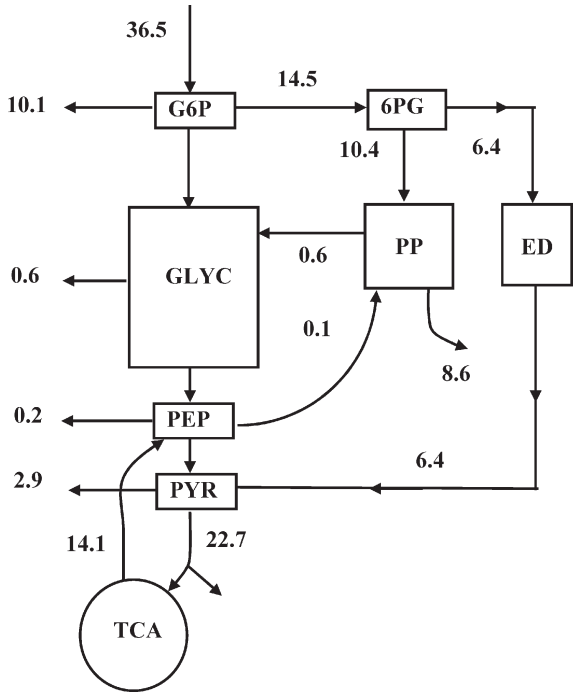
F i g . 5 e – *Simulation of the fluxes around phosphoenolpyruvate after the glucose impulse. The efflux flow are r_{PTS} (A black), r_{pk} (B gray), r_{PEPBM}(C green), and r_{DHAPS} (D purple), while the influxes are r_{GAPPEP} (E blue), and $r_{P,EPCK}$ (F red).*



F i g . 6 – *Experimentally observed response of oxaloacetate concentration to glucose impulse*

the influxes from GAP, r_{GAPPEP}, and from OxAc by r_{PEPCK}. The results show that the flux from OxAc is dominant from the very moment of the glucose impulse. At the same time, the flux from PEP to PYR stops and, even if temporarily, it reverses. The effluxes from PEP to DHAPS and biomass are relatively negligible in comparison with the main fluxes. The regulation of PEP is dynamically tied to the balance of OxAc by the TCA activity, being inferred from the experimental concentration of oxaloacetat shown in Fig. 6. Due to the consumption of PEP in PTS, with a time delay of 2.5 s after the glucose impulse, the concentration of OxAc presents a sharply increase in order to replenish the loss of PEP.

In Fig. 7. are presented the carbon integral mass flows in the *E. coli* cell (in mmol C_1 L^{-1}), dur-



F i g . 7 – *Integral mass balances of C_1 (mmol L^{-1}) evaluated during the 17 s under transient conditions upon glucose impulse*

ing the first 17 s after the glucose impulse. The flows are calculated by numerical integrating the reaction rates, and by using interpolated values of concentrations. Since time profiles of the reaction rates change drastically during the transient period, the calculated values are far from an equilibrium flux distribution. Also, due to the unsteady state of species concentrations, the balances do not sum up to zero. Hence, the differences are accounted by an increase or a decrease of the metabolite pools. The overall carbon balance is positive for (8.4 mmol L^{-1} C), being balanced by the CO_2 evolution and change in the pool concentration.

Conclusions

The main contributions of this work are related to testing the efficiency of three global optimisation algorithms for parameter estimation in complex kinetic models, and the essential improvements of the *E. coli* metabolic model with respect to regulation of PTS and to prediction of the oscillatory response of the C5 pool.

Modelling of complex enzymatic reactions systems is a difficult task. One of sources of the difficulty is related to the numerical optimization algorithms low efficiency to fit the data by minimizing the errors between model predictions and experimental measurements. In the present study, the non-gradient methods of Nelder-Mead, dynamic evolution and simulated annealing have been tested. These methods are considered as potentially global optimisation algorithms, by avoiding convergence toward local minima, usually present in the multi-variable/multi-parameter models. Nelder-Mead and DE-GA gave approximately similar convergence rates, but the DE could be recommended as being most suitable method due to its extensive random search in the parameter space. However, the efficiency of these methods is dependent on the selection of the initial parameter values, and of a number of "tune-up" factors, all making that the performance to be case dependent.

Application of the very accurate and stable integration of stiff ordinary differential kinetic equations, provided by software *Mathematica*,[10] essentially contributes to the quality and reliability of the model predictions.

As a result of the kinetic model modifications and of the efficient parameter estimation, a significant improvement in the overall prediction of the metabolite transients is reported. The obtained average standard deviation between the model and experimental data, of 0.4 mmol L^{-1}, is satisfactory. The model predicts a strong inhibition of the glucose transfer system by the accumulated G6P, oppositely to the effect of PEP depletion, as proposed by *Degenring* et al.[6] Most important, the model dynamically interrelates the balances of the glycolytic pool of PEP and PYR with the PTS and TCA. The result strong coupling of the mass balances of several pathways is an indication of a global concerted regulation of the whole sub-system. This sub-system represents in fact the main "traffic" controller of the catabolic cell processes and it possibly contains the main targets for genetic manipulations.

The model predictions of the oscillatory response in C5 pool are in agreement with other experimental observations, being a characteristic feature of two genotypes of *E. coli*.[12]

Based on the model simulations, dynamic flux balances around branching network metabolites are analysed. Mass distribution of carbon, i.e. C_1 equivalent metabolite, is determined by numerical integration of the model reaction rates. The overall mass balance does not tend to zero due to a significant change of concentration pools under the high drastic transient conditions imposed by the glucose impulse. The observed dynamic balances in the short time window of observations are basically related to the enzyme effects. However, the enzyme concentrations are a consequence of the physiological state of cells prior to the change in the environment.

The importance of the experimental method and the theoretical analysis of the dynamic metabolite balances has been proved by a successful application in the case of to anabolic pathways for aromatic synthesis.[12] As a general conclusion, modelling complex kinetics can be supported by a complementary application of the numerical flux control techniques and statistical estimation methods.[13]

In order to design further experimental and modelling investigations,[14] interpretation of dynamic metabolic balances is proved to be an essential aspect of the analysis. Such a metabolic flux balance analysis can be improved by using an impulse response technique applied to a synchronized cell population under a predefined and controlled metabolic state achievable in a chemostat.

ACKNOWLEDGEEMENT

Communication and assistance of D. Degenring is gratefully acknowledged.

Nomenclature

Metabolites

AcCoA – acetyl-coenzyme A
ADP – adenosindiphosphate
ATP – adenosintriphosphate
BPG – glycerate-1,3-bisphosphate

CIT – citrate
C5P – lumped pentose phosphate pool
DAHAP – 7-phospho-2-dehydro-3-deoxy-D-arabinoh--eptonate
DHAP- glycerine phosphate
ED – Entner-Doudoroff pathway
E4P – D-erythrose-4-phosphate
FBP – β-D-fructose-1,6-bisphosphate
F6P – β-D-fructose-6-phosphate
GAP – glyceraldehydes-3-phosphate
GTP – guanosine triphosphate
G3P – glycerol-3-phosphate
G6P – α-D-glucose-6-phosphate
NAD– diphosphopyridindinucleotide (oxidized)
NADH – diphosphopyridindinucleotide-phosphate (oxidized)
MUR – mureine
OxAc – oxaloacetat
PEP – phosphoenolpyruvate
PP – phospho-pentose pathway
2PG – glycerate-2-phosphate
3PG – glycerate-3-phosphate
6PG – 6-phospho-D-gluconate
PTS – phosphatransferase system
PYR – pyruvate
R5P – D-ribose-5-phosphate
Ri5P – D-ribulose-5-phosphate
S7P – D-sedoheptulose-7-phosphate
X5P – D-xylulose-5-phosphate

Enzymes

aldo – aldolase
dahps – dahp synthase
eno – enolase
gapdh – glyceraldehydes-3-phoshate dehydrogenase
g3pdh – glycerol-3-phoshate dehydrogenase
g6pdh – glycerol-6-phoshate dehydrogenase
pdh – pyruvate dehydrogenase
pepck – phosphoenolpyruavate carboxykinase
pfk – phosphofructokinase
pgdh – 6-phosphogluconate dehydrogenase
pgi – glucose-6-phosphate isomerase
pk – pyruvate kinase
pts – phosphotransferase system
ta – transaldolaze
tim – triose phosphate isomerase
tka – transketolase A
tkb – transketolase B

Variables

c – concentration, mol L^{-1}
E – energy, J
K – enzyme saturation constant, mmol L^{-1}
k – Boltzman constant
N – dimension
p – probability
r – intracellular concentration rate (flux), mol L^{-1} s^{-1}
S – sum of squares of residuals
T – temperature, K
t – time, s
V – maximum reaction rate, mol L^{-1} s^{-1}
v – reaction rate, mol L^{-1} s^{-1}
x – variable
z – variable

Greek

β – kinetic parameter
γ – scaling factor

Subscripts

a – random number
b – backward reaction
f – forward reaction
i – index
s – Michaelis-Menten saturation constant
m – measured data
u – input

Literature

1. *J. E. Bailey,* Science. **252** (1991) 1668.
2. *J. E. Bailey,* Biotechnol. Prog. **14** (1998) 8.
3. *U. Schaefer, W. Boos, R. Takors, D. Weuster-Botz,* Anal. Biochem. **270** (1999) 88.
4. *A. Buchholz, J. Hurlebaus, C. Wandrey, R.Takors,* Biomol. Eng. **19** (2002) 5.
5. *C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, M. Reuss,* Biotechnol. Bioeng. **79** (2002) 53.
6. *D. Degenring,* "Erstellung und Validierung mechanistischer Modelle für den mikrobiellen Stoffwechsel zur Auswertung von Substrat-Puls-Experimenten". Dissertation. University of Rostock, 2004.
7. *D. Degenring, C. Froemel, G. Dikta, R. Takors,* J. Process Control **14**(2004)729-745.
8. *S. Čerić,* "Dynamic modeling of *Escherichia coli* central metabolism response upon glucose impulse", Diploma work (in Croatian), University of Zagreb, 2005.
9. *G. Maria*, Chem. Biochem. Eng. Q., **18** (2004) 195.
10. *Wolfram Research* "*Mathematica*", v. 5.1 (2005).
11. *W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling,* "Numerical Recepies", Cambridge University Press, Cambridge, 1987.
12. *M. Oldiges, M. Kunze, D. Degenring, G. A. Sprenger, R. Takors,* Biotechol. Prog., **20** (2004) 1623.
13. *M. D. Haunschild, B. Freisleben, R. Takors, W. Wiechert,* Bioinformatics, **21** (8) (2005)1617.
14. *M. B. Ternbach, C. Bollman, C. Wandrey, R. Takors,* Biotechnol. Bioeng. **91** (2005) 356.